



(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

on: Thu Aug 21 09:17:50 1997; Masspar time 1415.67 Seconds

Tabular output not generated. 1184.648 Million cell updates/sec

Title: >US-08-469-637A-1  
Description: (1-1527) from US08469637A.seq  
Perfect Score: 1527  
N.A. Sequence: 1 GCCCCAGCCGCCGCCCTCCAA.....TTCAACTGCAAAAAAAA 1527  
Comp: GCGGGTCGCGCGGAGGTT.....AAGTTGACCTTTTITTTTT

Scoring table:  
TABLE default  
Gap 6

Mismatch STD : Dbase 0; Query 0  
362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1.new3  
1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV  
9: ORG 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC  
17: VIR

Database:

genbank99  
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7  
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2  
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3  
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10  
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3  
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG  
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7  
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2  
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9  
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15  
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7  
92: ROD8 93: STR 94: STR 95: UNA 96: VRL1 97: VRL2 98: VRL3  
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9  
105: VRL10

Database:

genbank.new3  
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV  
112: MAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2  
118: ROD 119: SYN 120: UNA 121: VRL  
u-emb150.99  
122: part1

Database:

Statistics: Mean 11.533; Variance 4.446; scale 2.594

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
C 1	32	2.1	215	57	128278	Sequence 5 from patent	4.58e-06
C 2	30	2.0	215	57	128278	Sequence 5 from patent	1.31e-04
C 3	28	1.8	354	111	OF089259	Oxytricha fallax 57KD	3.33e-03
C 4	28	1.8	354	8	OF089259	Oxytricha fallax 57KD	3.33e-03
C 5	25	1.6	354	111	OF089259	Oxytricha fallax 57KD	3.26e-01
C 6	25	1.6	354	8	OF089259	Oxytricha fallax 57KD	3.26e-01
C 7	25	1.6	1441	51	S66477	glial fibrillary acid	3.26e-01
C 8	25	1.6	1933	92	CRAE1	Carassius auratus (GF	3.26e-01
C 9	24	1.6	7095	99	RRMAB15	R. norvegicus mRNA for	1.38e+00
C 10	24	1.6	7322	64	PSCHSAB	P. setium gene for ch	1.38e+00
C 11	24	1.6	16632	35	CEC29F3	Caenorhabditis elegans	1.38e+00
C 12	24	1.6	47745	38	CEC09D4	Caenorhabditis elegans	1.38e+00
C 13	24	1.6	115419	34	HS199F10	Human DNA sequence **	1.38e+00
C 14	24	1.6	150894	20	D90904	Synechocystis sp. PCC	1.38e+00
C 15	23	1.5	1344	20	ECACGA	E. coli acea gene cod	5.59e+00
C 16	23	1.5	1747	21	ECACGA	Escherichia coli 180C	5.59e+00
C 17	23	1.5	1851	117	HS19PA4	H. sapiens LIPA gene,	5.59e+00
C 18	23	1.5	1851	75	HS19PA4	H. sapiens LIPA gene,	5.59e+00
C 19	23	1.5	2358	21	EC01DHRPA	E. coli acea gene enco	5.59e+00
C 20	23	1.5	3021	69	YSCGLN3	S. cerevisiae nitrogen	5.59e+00
C 21	23	1.5	3155	122	A31763	DNA expression cassette	5.59e+00
C 22	23	1.5	3177	43	LSRNBGRP	L. stagnalis mRNA for	5.59e+00
C 23	23	1.5	3720	20	ECACB	E. coli genes acea an	5.59e+00
C 24	23	1.5	6000	122	A31767	DNA expression cassette	5.59e+00
C 25	23	1.5	14833	31	ECAE000474	Escherichia coli from	5.59e+00
C 26	23	1.5	38586	37	CEFE55C12	Caenorhabditis elegans	5.59e+00
C 27	23	1.5	43100	65	SCE9379	Saccharomyces cerevis	5.59e+00
C 28	23	1.5	145750	33	HS11A724	Human DNA sequence **	5.59e+00
C 29	23	1.5	176195	22	ECOW89	E. coli chromosomal r	5.59e+00
C 30	22	1.4	618	108	FR0002420	F. rubripes GSS sequen	2.15e+01
C 31	22	1.4	618	3	FR0002420	F. rubripes GSS sequen	2.15e+01
C 32	22	1.4	1230	67	SLU39443	Lycopodium esculent	2.15e+01
C 33	22	1.4	1245	61	CREPHOSRI	C. reinhardtii phospho	2.15e+01
C 34	22	1.4	1256	47	OCSPI7GN	O. cuniculus SP17 gene	2.15e+01
C 35	22	1.4	1761	67	SCYLO085C	S. cerevisiae fus2 gen	2.15e+01
C 36	22	1.4	2492	65	SCDNAMEFUS2	S. cerevisiae fus2 gen	2.15e+01
C 37	22	1.4	3699	67	SCYLO084W	S. cerevisiae chromoso	2.15e+01
C 38	22	1.4	3796	90	MUSTNFR1	Murine tumor necrosis	2.15e+01
C 39	22	1.4	17013	20	EAAMST	E. amylovora (Ea7/74)	2.15e+01
C 40	22	1.4	40267	77	HSU159B9	Human DNA sequence fr	2.15e+01
C 41	22	1.4	90468	34	HS465610	Human DNA sequence **	2.15e+01
C 42	22	1.4	99370	115	ATAC000103	Genomic sequence for	2.15e+01
C 43	22	1.4	99370	5	ATAC103	Genomic sequence for	2.15e+01
C 44	22	1.4	99370	32	ATAC000103	*** SEQUENCING IN PRO	2.15e+01
C 45	22	1.4	257912	103	PBU42580	Parametium bursaria C	2.15e+01

## ALIGNMENTS

RESULT 1  
LOCUS 128278 215 bp DNA PAT 30-OCT-1996  
DEFINITION Sequence 5 from patent US 5569830.  
ACCESSION 128278  
NID g1819054

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

1 (bases 1 to 215)  
Bennett, A., Labavitch, J.M., Powell, A. and Stotz, H.  
Plant inhibitors of fungal polygalacturonases and their use to control fungal disease  
Patent: US 5569830-A 5 29-OCT-1996;

FEATURES

source

BASE COUNT

ORIGIN

Location/Qualifiers

1..215

/organism="unknown"

15 a 8 c 25 g 26 t 141 others

[illegible][illegible]

RA Doak T.G., Williams K., Witherspoon D.J., Herrick G.;  
RT Submitted (11-FEB-1997) to the EMBL/GenBank/DBJ databases.  
RL Oncological Science, University of Utah, School of Med. Rm5C334,  
RL USA, UT 84132, USA  
FH Key Location/Qualifiers  
FT 1.354  
FT /organism="Oxytricha fallax"  
FT /strain="9D1"  
FT /transposon="TBE1"  
FT /note="this is a bulk sequence that was generated from a  
FT PCR product that represents many transposon templates"  
FT <1..>354  
FT /codon\_start=1  
FT /product="57kd zinc finger/protein chimera"  
FT /transl\_table=6  
FT /db\_xref="PID:g1881676"  
FT /translation="HTRDLKHLKAHKKKXEXXXXXLKKLRKAREXXXXXX  
FT AXEXXXXXXRNXLXSEXTKIMIKIQYKRIPLVAQIDLTSLQSYLILEDSEFDKRVIX  
FT HOYAKVYFENVLX"  
D  
FT Sequence 354 BP: 106 A; 42 C; 41 G; 54 T; 111 others;  
FT  
FT Query Match 1.6%; Score 25; DB 111; Length 354;  
FT Best Local Similarity 23.7%; Pred. No. 3.26e-01;  
FT Matches 23; Conservative 38; Mismatches 36; Indels 0; Gaps 0;  
DB 110 hhyhbwargbwydygcykaayaagcwgmgarwyaswrtgtaar 158  
DB 802 CTGGTGAAGTGTGGAACATCAAAACAAGACCAAGA-TATATGTCAGAGATCATCA 860  
DB 170 accwyttrcmctcdgagrtmacd 192  
DB 861 AGATATGACCTCTGTGAAACA 883  
OY 861 AGATATGACCTCTGTGAAACA 883  
RESULT 5 OFU89259 354 bp DNA INV 14-MAR-1997  
LOCUS Oxytricha fallax 57kd zinc finger/protein chimera gene, partial  
DEFINITION cds.  
ACCESSION U89259  
NID g1881675  
KEYWORDS  
SOURCE  
ORGANISM Oxytricha fallax.  
Oxytricha fallax  
Eukaryotes; mitochondrial eukaryotes; Alveolata; Ciliophora;  
hypotrichs; Stichotrichida; Oxytricha.  
REFERENCE  
AUTHORS 1 (bases 1 to 354)  
Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G.  
TITLE A proposed superfamily of transposase genes: transposon-1like  
elements in ciliated protozoa and a common 'D35E' motif  
Proc Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)  
JOURNAL 94134747  
MEDLINE  
REFERENCE 2 (bases 1 to 354)  
Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.  
TITLE Selection on the protein-coding genes of the TBE1 family of  
transposable elements in the ciliates Oxytricha fallax and O.  
trifallax  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 354)  
Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.  
TITLE Direct Submission  
JOURNAL Submitted (11-FEB-1997) Oncological Science, University of Utah,  
School of Med. Rm5C334, USA, UT 84132, USA  
FEATURES  
source  
1..354  
/organism="Oxytricha fallax"  
/strain="9D1"  
/transposon="TBE1"  
/note="this is a bulk sequence that was generated from a  
PCR product that represents many transposon templates"

CDS  
1..>354  
/codon\_start=1  
/product="57kd zinc finger/protein chimera"  
/db\_xref="PID:g1881676"  
/transl\_table=6  
/translation="HTRDLKHLKAHKKKXEXXXXXLKKLRKAREXXXXXX  
OXEXXXXXXRNXLXSEXTKIMIKIQYKRIPLVAQIDLTSLQSYLILEDSEFDKRVIX  
XDQYAKVYFENVL"  
BASE COUNT 106 a 42 c 41 g 54 t 111 others  
ORIGIN  
FT  
FT Query Match 1.6%; Score 25; DB 111; Length 354;  
FT Best Local Similarity 23.7%; Pred. No. 3.26e-01;  
FT Matches 23; Conservative 38; Mismatches 36; Indels 0; Gaps 0;  
DB 99 rycdgrgaryhhyhbwargbwydygcykaayaagcwgmgarwyaswrtgtaar 158  
DB 799 GGAAGTGTGTCTGTGAGCTGTGCGTGTATTCCTCTACACTCTGCGTTTA 740  
DB 159 gawdrcrcgaaccwyttrcmctcdgagrtmacdaar 195  
DB 739 CTGTGTCGCGCAAGTGTCTACCAAGACACTTAAG 703  
CP  
RESULT 6 OFU89259 standard: DNA; INV; 354 BP.  
ID OFU89259  
AC U89259;  
NI g1881675  
DT 13-MAR-1997 (Rel. 51, Created)  
DT 13-MAR-1997 (Rel. 51, Last updated, Version 1)  
DE Oxytricha fallax 57kd zinc finger/protein chimera gene, partial  
DE cds.  
OS Oxytricha fallax  
OC Eukaryotes; mitochondrial eukaryotes; Alveolata; Ciliophora;  
OC hypotrichs; Stichotrichida; Oxytrichidae; Oxytricha.  
RN [1]  
RP 1-354  
RX MEDLINE; 94134747.  
RA Doak T.G., Doerder F.P., Jahn C.L., Herrick G.;  
RT "A proposed superfamily of transposase genes: transposon-1like  
RT elements in ciliated protozoa and a common 'D35E' motif";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:942-946(1994).  
RN [2]  
RN 1-354  
RA Witherspoon D.J., Doak T.G., Williams K., Seger J., Herrick G.;  
RT "selection on the protein-coding genes of the TBE1 family of  
RT transposable elements in the ciliates Oxytricha fallax and O.  
RT trifallax";  
RL unpublished.  
RN [3]  
RN 1-354  
RA Doak T.G., Williams K., Witherspoon D.J., Herrick G.;  
RT Submitted (11-FEB-1997) to the EMBL/GenBank/DBJ databases.  
RL Oncological Science, University of Utah, School of Med. Rm5C334,  
RL USA, UT 84132, USA  
FH Key Location/Qualifiers  
FT 1.354  
FT /organism="Oxytricha fallax"  
FT /strain="9D1"  
FT /transposon="TBE1"  
FT /note="this is a bulk sequence that was generated from a  
FT PCR product that represents many transposon templates"  
FT <1..>354  
FT /codon\_start=1  
FT /product="57kd zinc finger/protein chimera"  
FT /transl\_table=6  
FT /db\_xref="PID:g1881676"  
FT /translation="HTRDLKHLKAHKKKXEXXXXXLKKLRKAREXXXXXX  
FT AXEXXXXXXRNXLXSEXTKIMIKIQYKRIPLVAQIDLTSLQSYLILEDSEFDKRVIX  
FT O



```

/tissue_type="brain"
/cell_type="C6 glioma"
BASE COUNT      2124 a      1836 c      1799 g      1316 t
ORIGIN
Query Match      1.6%; Score 24; DB 92; Length 7095;
Best Local Similarity 76.1%; Pred. No. 1.38e+00;
Matches 35; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db      1004 gcaaggaatcagattcatgcagcagtgagtggaaccaaca 1049
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      305 GCAAGGAGCTCAGTACGTCAAGCAGAGTGCATCGACCCACAA 350

RESULT 10
LOCUS      PSCHSAB      7322 bp      DNA
DEFINITION P.sativum gene for chalcone synthase.
ACCESSION  X80007
            g510542
ORFS
chalcone synthase; CHS-1A gene; CHS-1B gene; class 1 gene.
pea.
ORGANISM    Pisum sativum
Eukaryote; mitochondrial eukaryotes; Viridiplantae;
Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
Magnoliopsida; Rutanae; Sapindales; Fabaceae; Papilionoideae;
Pisum.
REFERENCE   1 (bases 1 to 7322)
AUTHORS     Hellens, R.P.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 7322)
AUTHORS     Hellens, R.P.
TITLE       Direct Submission
SUBMITTED   (01-JUL-1994) R.P. Hellens, John Innes Institute, Colney
Lane, Norwich NR4 7UH, UK
FEATURES
     source          location/qualifiers
                     1..7322
                     /organism="Pisum sativum"
                     /strain="J1 813"
                     /clone_lib="labda GEM-11"
                     /clone="B"
                     /map="linkage group 3"
                     <1040..1217
                     /gene="CHS-1A"
                     /number=1
                     join(<1040..1217,1328..>2319)
                     /gene="CHS-1A"
                     join(1040..1217,1328..2319)
                     /gene="CHS-1A"
                     /EC_number="2.3.1.74"
                     /codon_start=1
                     /product="naringenin-chalcone synthase"
                     /db_xref="PID:g510543"
                     /translation="MTVNEIRROARAGPATVATGATAPNOCVDSOSTYDPYFRIT
                     NSHRTLEKRFQRMCDKSKIKRYMHLTEILKNPSLCEYAPSLDARQDVVVEV
                     PKLGEKATVAKIEWGQPKSKITLHICTTSGVDMPPADQTLKLLGLRYVARYMY
                     QQCFAGGTGLRLAKDLAENNGARVLVWSEITAVTFRQSPDTHDSLGLQALFGDG
                     AAAYIVGSDPLPOVEKPLFELVWTAQTIIVPDSGALDGLHREAGLTFHLKDVPSLVS
                     KNIEKALVEAFQPLNISIDYNSIFWIAHPGGPALIDOVEAKLGLKORKMOATRHVLSY
                     GNMSSACVLFILDEMRKSKEDLATGTGELGVLFQFGPLTVEIVLHSHAT"
                     1218..1327
                     /gene="CHS-1A"
                     /number=1
                     1328..>2319
                     /gene="CHS-1A"
                     /number=2
                     join(<4336..4513,4642..>5633)
                     /gene="CHS-1B"
                     join(4336..4513,4642..5633)
                     /gene="CHS-1B"
                     /EC_number="2.3.1.74"
                     /codon_start=1
                     /product="naringenin-chalcone synthase"
                     /db_xref="PID:g510544"

```

```

/translation="MTVDEIRROAKRAGPATVLAIGTATPNCVDSOSTYDPYFRIT
NSEHRTLEKRFQRMCDKSKIKRYMHLTEILKNPSLCEYAPSLDARQDVVVEV
PRLGKEATVAKIEWGQPKSKITLHICTTSGVDMPPADQTLKLLGLRPHYRYMY
QOQCFAGGTGLRLAKDLAENNGARVLVWSEITAVTFRQSPDTHDSLGLQALFGDG
AAAYIVGSDPLPOVEKPLFELVWTAQTIIVPDSGALDGLHREAGLTFHLKDVPSLVS
KNIEKALVEAFQPLNISIDYNSIFWIAHPGGPALIDOVEAKLGLKORKMOATRHVLSY
GNMSSACVLFILDEMRKSKEDLATGTGELGVLFQFGPLTVEIVLHSHAT"
<4336..4513
/gene="CHS-1B"
/number=1
4514..4641
/gene="CHS-1B"
/number=1
4642..>5633
/gene="CHS-1B"
/number=2
BASE COUNT      2350 a      1307 c      1264 g      2401 t
ORIGIN
Query Match      1.6%; Score 24; DB 64; Length 7322;
Best Local Similarity 81.6%; Pred. No. 1.38e+00;
Matches 31; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db      654 aaagtgtaacctctatggaataaccaccaatggt 691
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      1404 AAAGACTAATCATCTCTCCATTAACCCCAATGCTT 1441

RESULT 11
LOCUS      CEC29F3      18632 bp      DNA
DEFINITION Caenorhabditis elegans cosmid C29F3.
ACCESSION  Z81043
NID        91627634
KEYWORDS   Caenorhabditis elegans.
SOURCE     Eukaryote; mitochondrial eukaryotes; Metazoa; Nematoda;
ORGANISM    Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditioidea;
             Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 18632)
AUTHORS     Matthews, L.
JOURNAL     Direct Submission
SUBMITTED   (21-OCT-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jesse.sanger@ac.uk or tw@nemastock.wustl.edu
2 (bases 1 to 18632)
Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurtry, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Ritken, L., Roopra, A.,
Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, D., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Spratt, J. and Woldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
Nature 368 (6466), 32-38 (1994)
94150718
JOURNAL MEDLINE
COMMENT
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
clone C29F3. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green, ms in preparation), and other

```



REFERENCE 1 (bases 1 to 115419)  
Eukaryotes: mitochondrial eukaryotes: Metazoa: Chordata: Vertebrata: Eutheria: Primates: Catarrhini: Hominiidae: Homo. Hunt, A.  
Direct Submission  
Submitted (06-NOV-1996) Wellcome Trust Genome Camps, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hummer@anger.ac.uk  
Cloned from: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished sequence: BK799F10 Contig\_ID: 00052 Length: 992 bp Unfinished sequence: BK799F10 Contig\_ID: 00061 Length: 885 bp Unfinished sequence: BK799F10 Contig\_ID: 00185 Length: 12614 bp Unfinished sequence: BK799F10 Contig\_ID: 00949 Length: 830 bp Unfinished sequence: BK799F10 Contig\_ID: 00953 Length: 785 bp Unfinished sequence: BK799F10 Contig\_ID: 00960 Length: 958 bp Unfinished sequence: BK799F10 Contig\_ID: 01506 Length: 6932 bp Unfinished sequence: BK799F10 Contig\_ID: 01718 Length: 38510 bp Unfinished sequence: BK799F10 Contig\_ID: 02165 Length: 15742 bp Unfinished sequence: BK799F10 Contig\_ID: 02195 Length: 2254 bp Unfinished sequence: BK799F10 Contig\_ID: 02223 Length: 1000 bp Unfinished sequence: BK799F10 Contig\_ID: 02224 Length: 1000 bp Unfinished sequence: BK799F10 Contig\_ID: 02225 Length: 1341 bp Unfinished sequence: BK799F10 Contig\_ID: 02226 Length: 1016 bp Unfinished sequence: BK799F10 Contig\_ID: 02214 Length: 19570 bp.  
\*\*\*  
\*\*\* WARNING: Phase 1 High Throughput Genome Sequence \*\*\*  
\*\*\*  
\* This sequence is unfinished. When sequencing is complete, \* the sequence data presented in this record will be replaced \* by a single finished sequence with the same accession number.  
Location/Qualifiers  
source 1..115419  
/organism="Homo sapiens"  
/clone="799F10"  
/chromosome="22"  
BASE COUNT 24169 a 28425 c 26764 g 23990 t 12071 others  
ORIGIN  
Query Match 1.68; Score 24; DB 34; Length 115419;  
set Local Similarity 80.0%; Pred. No. 1.38e+00;  
atches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Db 109739 tcttctgttctgttcttcttcacactctgcagctgcaaga 109778  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
CP 833 TCTTTGTTTGATGTTTCCATCACTTCAGCAGCTGGAAG 794  
RESULT 14  
LOCUS D90904 150894 bp DNA BCT 20-NOV-1996  
DEFINITION Synecocystis sp. PCC6803 complete genome, 6/27, 630555-781448.  
ACCESSION D90904  
NID g1652225  
KEYWORDS formaldehyde dehydrogenase; ABC transporter; DNA ligase; DNA polymerase III alpha subunit; Gumb protein; LysR transcriptional regulator; Mg-protoporphyrin IX; N-acetylornithine aminotransferase; NADH dehydrogenase subunit 4; NADH dehydrogenase subunit 5; acetate kinase; aspartoacylase(ASP); cation-transporting ATPase; cell division protein FtsH; chemotaxis protein CheB; cytochrome oxidase d subunit I; cytochrome oxidase d subunit II; diene lactone hydrolase; dihydroflavonol 4-reductase; dihydropterocate pyrophosphorylase; esterase; ferrous iron transport protein B; fibrillin; formaldehyde dehydrogenase (glutathione); high light-inducible protein; leader peptidase I; lysostaphin; malonyl coenzyme A-acyl carrier protein transacylase; methyl-accepting chemotaxis protein (MCP); oligopeptide transport ATP-binding

protein Opp; oxygen independent coprophorhyllinogen III oxidase; penicillin-binding protein 1b; phenoxypybenzoate dioxygenase; phycoerythrin a subunit; phycoerythrin associated linker protein; phycoerythrin b subunit; potassium channel; protein conferring resistance to acetazolamide, zam; regulatory component of sensory transduction system; regulatory components of sensory transduction system; ribonuclease II; sensory transduction histidine kinase; serine esterase; serine/threonine protein kinase; seryl-L-tryptophan synthetase; sporulation protein SpoIID; sulfolipid biosynthesis protein SdgB; tRNA-Arg; tRNA-Leu; tyrosyl tRNA synthetase. Synecocystis sp. (strain:PCC6803) DNA.  
SOURCE Synecocystis sp.  
ORGANISM Synecocystis sp.  
REFERENCE 1 (bases 1 to 150894)  
Eubacteria: Cyanobacteria: Chroococcales: Synecocystis. Tabata, S.  
Direct Submission  
Submitted (28-JUN-1996) to the DDBJ/EMBL/Genbank databases. Satoshii Tabata, Kazusa DNA Research Institute, Laboratory of Gene Structure 2, 152-3, Tananuchino, Kisarazu, Chiba 292, Japan (E-mail: tabata@kazusa.or.jp, Tel: +81-438-52-3933, Fax: +81-438-52-3934)  
2 (sites)  
Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosewa, M., Sugiyura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naito, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S.  
Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions  
DNA Res. 3 (3), 109-136 (1996)  
97061201  
JOURNAL MEDLINE  
COMMENT  
POTENTIAL protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.  
Location/Qualifiers  
source 1..150894  
/organism="Synecocystis sp."  
/strain="PCC6803"  
complement(1..1038)  
/note="ORF\_ID: s110998"  
/codon\_start=1  
/product="LysR transcriptional regulator"  
/db\_xref="PID:g1652226"  
/transl\_table=1  
/translation="MSDIPFTLDRLIKKAIASGSEFKRAADRLTYSOPAVSLQVNTL EKOLSVPLFDRGKRAQLTEAGHLLNLYGKIIITLCOTCRALDNLQGLTIYVA SOTGTLLPLPMIGMRQCPDYVQVQVSTRTANGVANGVDLAIIGSEVPALQ ETIVLVLYAEDELALIPVLPPLAOETIOEDLYKLFISLDSOSTIRKVIDVLSO GEIDTRKLIKEMELNSIEAIKNAVQSLGAFTVTAIEKLENNVLAHPIKRVETIR VLSVILNPNTYRSKASAPFIREVLPOFSTHPDALDEPRLFANPSSNNDROGDGD GKSIEIDSVY"  
complement(1342..1539)  
/note="ORF\_ID: s11911"  
/codon\_start=1  
/product="hypotheical protein"  
/db\_xref="PID:g1652227"  
/transl\_table=1  
/translation="MSTQOQARALMRHGFIRKROQSMLSRAAEIGVNAEDFWTT VQGRPOSSFRRTYRSNALS"  
1659..2603  
/note="ORF\_ID: s11019"  
/codon\_start=1  
/product="hypotheical protein"  
/db\_xref="PID:g1652228"  
/transl\_table=1  
/translation="MENSVPYPMRYGTYLDTVDFTDQLFGGNPLAVPDAGLTDGQNO KTAIEINISERVFLPVPTETGKGRLEITPKRLADAGHPTGTITLLDLPSPSL VITWIOEPEVGLPVTLHYKGOVOTETVPAOLDKASPCASCDLALLSLISIDQ LQOEYEPQASCGLPFLFPLINEALNRSFSPWQNLGQVADYCLAPGP SLGSDPKLILHGRMFARGLIADBPATGSGVAALGVLDLRDTPGSHHQIIOGKAL GRPQDILTYVKDQGGIRAVKAGRSVLVSEGFNNLGN"  
2750..2830  
/product="tRNA-Leu"



CDS  
 complement(3325..4326)  
 /note="ORF\_ID:s110997"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /db\_xref="PID:g1652229"  
 /transl\_table=11  
 /translation="MAPKSHCHNGDKMKNORSNGEMTAVAPYOSFHHIGLIGLALASV  
 PLSACITTPSPFPAVONQVNDROEDVDAONSDDTYTETEDPNSDERKLGKVBK  
 DVASASISYQEKHDGSEWPEVTYKFKNKIVATLGSSPMFIALQITDDNDROPY  
 PALEFATYGGACNCEVETSNQNSMSVDFGFGNGPAPADLNDNDVEVE  
 VDRFATYSSYGAAPAOIHALONGQVAVFSEFPOFIHRENAOSKENDLPETVA  
 ODEKNGFLAAYVANKALIGELDEGQTMKXYDRESMDGLTNCLEYDDQSNLNBVK  
 YDSYPAALAFIYAGTYEPAKFEAENNTL"  
 complement(4338..5693)  
 /note="ORF\_ID:s110996"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /db\_xref="PID:g1652230"  
 /transl\_table=11  
 /translation="MNVSHRRYHITFGCCOMKADSEFMAGLIENTGMTYTDPPNOAD  
 LVYNTQSLRDNAQKVSYLGRQARKQVEPELTLVAGVAAQOGEQLRRVEPLD  
 LVMGPANRLDOLLEOVAGSOVAVTESLHIMEDITKPRRESTVSAWNIYYGNER  
 CSCVVPNVGVSOSRTPALYGEVEMLAQOQFEVTLGQNDAYGRDLPGTTPSGR  
 HLHTLDLYHVDIGIDRLRATSHPRYFTEFLIOACOELPKVCEHFHIDPOSDN  
 DILAKKRGYTRKRYQIIEKIRKVPDAISADYVGFPGTEAFQENTLNLIEBVG  
 PDLNTAASPRGTPAEMNDOLSEVAGDRQLQNLHLVSTQAMERSORTIGRVEV  
 LVGEMLKSPQGYMGRTRGNRLTFQGEISELGLKTVPVKITEARFSLTGALSIVT  
 A"  
 5818..6969  
 /note="sgdB"  
 /note="ORF\_ID:s1r1020"  
 /codon\_start=1  
 /product="sulfolipid biosynthesis protein SqdB"  
 /db\_xref="PID:g1652231"  
 /transl\_table=11  
 /translation="MRAIVIGDGYCGMATAFLYSKGYEVLGDSLVARYMDAOLGA  
 ERTLPAPLRORLDRWELTGKIDLFIDINDYPLTNALRQFODPAVHGEQNSA  
 PSMIDREHVALTOANNVGNLNLALYALHEDPDLVNLGTGEGITNIDIEBGT  
 TTEHKGKDTLPYKQSGSYHLKSVHDSHNFHACKINGLRATDINOIVGVLTET  
 TGDEMELINRLDVGFTALNRFICQAAIGHFLVYVGGQTRGLIDRDVRCLEL  
 AINPADKQOFRENOYTELFESVGLDQAOVKAGADLGLKVEIDHLENRVELEEHYF  
 NAVNTLGLDGLPHFLSDLSLDSLNFATKRYKDRDQKHILPKYTWRC"  
 7112..7402  
 /note="ORF\_ID:s1r1698"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /db\_xref="PID:g1652232"  
 /transl\_table=11  
 /translation="MADPLTPAISDRICKHMDHSAIATLYAOUFGOOTDTYMAQMO  
 AIDPTMDLVESGSGSKTIRITFEOPPLKDSEDAHOVLIAAKQARSQKNS"  
 complement(7439..8134)  
 /note="ORF\_ID:s110995"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /db\_xref="PID:g1652233"  
 /transl\_table=11  
 /translation="MTVALDREIIPYSDGQPMADNTEQFEMIVLKENLECLFAHNP  
 DVEFGDLMVPEGHREIRVADVVAALGRPKGSGSYROMQENNAQOVVEILISP  
 GNLKEMTKKLFKFDHGVGEYVYDPPDNELTGLRIGELTIIEMAHMWSPLGI  
 KFELSAETLRVYYPDGRPLSTVALATQAEQASQRANEAQRAERERKLAELAEEND  
 RKLALAEAGIDV"  
 complement(8220..9029)  
 /note="ORF\_ID:s110994"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /db\_xref="PID:g1652234"  
 /transl\_table=11  
 /translation="MSTNFFVPSIOHGVCCQKTIYSINTVIMFLANLYTWLLG  
 LGLLCGLMTISGDGDEDDVENSLEASSDDITHLDVLDGDMDEBEAVALMAYQV  
 LSEFGIGKVPMLILLIDISLWGVIGMIVNIVATVGTMPRLMMAOIIIFLSLAI  
 SLWIGRLASPIALHLFKTFSDVSTERYIGCTGTYSKRLPYLANGTIGOAIVYNAG  
 NLITIVSLPADVATVPHNQEILIDQSPKGYGLAIKAKSDSDDEKMLKS"  
 complement(8962..10059)

CDS  
 /note="ORF\_ID:s110993"  
 /codon\_start=1  
 /product="potassium channel"  
 /db\_xref="PID:g1652235"  
 /transl\_table=11  
 /translation="MGISGSOEMLNLIDROGRRLROELMAGAITLACLFVCTAVY  
 RYVEMTLDLFYMTITTLATVGFGEHPHPSPARSLRTIILILGLLITGMVRPTE  
 AFIOGYFODLSRRROEQKAVIERLADHYLLCGYGRGOIAEFVEMNPFVYDIASPE  
 VIOKLKLDVAVLQGDALDEILLAHIERAICIVSALSDEANLYVLSAKTINPR1  
 PAIARASSEAVOKLRAGADEVVSPTTGCRRLAALRQVVSFVGDILTGADRSF  
 YMEERISAECPITGOTLREAOLRQASGALILAIRQDRKLIYGPMDTLLDADSL  
 ICLGVDELRALNQLCLINPARVRLRPNRH"  
 10183..11472  
 /note="argD"  
 /note="ORF\_ID:s1r1022"  
 /codon\_start=1  
 /product="N-acetylornithine aminotransferase"  
 /db\_xref="PID:g1652236"  
 /transl\_table=11  
 /translation="MTSPVSVESVDAQFAVTDISPAEKTADFTYVANTGRPPI  
 AIAHQGSTLMDTEGSKSLDFAVAGIATCLGHAPALVAVASDQIQRLHVSNTLYIP  
 EOGELAKIVHSCADRVFENSGAEANEAIKLVKRYAHTVLDLPQVLYLTKASF  
 HGRATATTATGQPKYQOQFEDPLVPGDYVPYNDIRSLKENVADLDGNSVVAIFPE  
 PLOGEGVBPDLAVFKRREICDONDILVDFEVOVQVGRGTLMGTEHGVDPDIF  
 TSAKGLAGVYIGAMCKKFCEDVPEPCHASTPFGNPLACAAGLAVLKTIDGRDLN  
 VQAGEQURSGLAIEIKNOYPTLFEVRGWLINGLETSAESSLTSVEIVKAAEKGGL  
 LAPAPRYLRVPLVYVTEAIEAIVELRQAIATLV"  
 11599..11913  
 /note="ORF\_ID:s1r1023"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /db\_xref="PID:g1652237"  
 /transl\_table=11  
 /translation="MKNLMOQLPDRLPDQEDFLLENKNTLIERIISTGQITPGEVY  
 DQDRREWVLLQGAKEIYDYGAVISLGGQGYLLIPRQKRRVYFTTPEPCILALH  
 FS"  
 11920..12561

CDS  
 Note: remainder of annotations omitted.  
 Query Match  
 Best Local Similarity 81.6%; Score 24; DB 20; Length 150894;  
 Matches 31; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 4235 aaactcgtatcgtgacatcactcactccccc 4272  
 1111 1111 1111 1111 1111 1111  
 QY 1388 AACTCATGATGTGAGAAAGACTAACATCTCTCCAA 1425

RESULT 15  
 LOCUS ECACEA 1344 bp DNA BCT 12-SEP-1993  
 DEFINITION E. coli aceA gene coding for isocitrate lyase (EC 4.1.3.1).  
 ACCESSION X07543  
 NID 940884  
 KEYWORDS aceA gene; isocitrate lyase; isocitrate lyase.  
 SOURCE Escherichia coli.  
 ORGANISM Escherichia coli.  
 Euacterichia: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.  
 1 (bases 1 to 1344)  
 Cozone,A.J.  
 Direct Submission  
 Submitted (06-MAY-1988) Cozone A.J., University of Lyon,  
 Laboratory of Molecular Biology, 43 Boulevard du onze Novembre,  
 69632 Villeurbanne, France  
 2 (bases 1 to 1344)  
 Riell,C., Bleicher,F., Duclos,B., Cortay,J.C. and Cozone,A.J.  
 Nucleotide sequence of the aceA gene coding for isocitrate lyase in  
 Escherichia coli  
 Nucleic Acids Res. 16 (12), 5689 (1988)  
 88262573  
 JOURNAL MEDLINE  
 FEATURES  
 source Location/Qualifiers  
 1..1344



